Measurement Error, Biases, and the Validation of Complex Models for Blood Lead Levels in Children

Raymond J. Carroll[1] and Christian D. Galindo[2]

[1]Nutrition and Toxicology, Department of Statistics, Texas A&M University, College Station, Texas; [2]Genentech, Inc., San Francisco, California

Measurement error causes biases in regression fits. If one could accurately measure exposure to environmental lead media, the line obtained would differ in important ways from the line obtained when one measures exposure with error. The effects of measurement error vary from study to study. It is dangerous to take measurement error corrections derived from one study and apply them to data from entirely different studies or populations. Measurement error can falsely invalidate a correct (complex mechanistic) model. If one builds a model such as the integrated exposure uptake biokinetic model carefully, using essentially error-free lead exposure data, and applies this model in a different data set with error-prone exposures, the complex mechanistic model will almost certainly do a poor job of prediction, especially of extremes. Although mean blood lead levels from such a process may be accurately predicted, in most cases one would expect serious underestimates or overestimates of the proportion of the population whose blood lead level exceeds certain standards. — *Environ Health Perspect* 106(Suppl 6):1535–1539 (1998). *http://ehpnet1.niehs.nih.gov/docs/1998/Suppl-6/1535-1539carroll/abstract.html*

Overview

This article focuses on the role of measurement error in linear regression, with special emphasis on relating a single environmental lead media exposure to blood lead in children and validating complex mechanistic models. For further details and a general overview of the topic, including work on multiple media exposures and non-linear regression models, see Carroll et al. (*1*); Fuller (*2*) should also be consulted for the linear model. Our emphasis on the simplest model is deliberate because the ideas are best understood in this context. However, all of the global conclusions carry over to more complex models.

There are three major points in this article:

- Measurement error causes biases in regression fits. If one could accurately measure exposure to the environmental lead media, the line obtained would differ in important ways from the line obtained when one measures exposure with error.
- The effects of measurement error vary from study to study. It is dangerous to take measurement error corrections derived from one study and apply them to data from entirely different studies or populations.
- Measurement error can falsely invalidate a correct (complex mechanistic) model. If one builds a model such as the integrated exposure uptake biokinetic (IEUBK) model carefully, using essentially error-free lead exposure data, and applies this model to a different data set with error-prone lead exposures, the complex mechanistic model will almost certainly do a poor job of prediction, especially of extremes. Although mean blood lead levels from such a process may be accurately predicted, in most cases one would expect serious underestimates or overestimates of the proportion of the population whose blood lead level exceeds certain standards.

Measurement Error Models

Measurement error models have a common structure:

- An underlying model for a response (e.g., blood lead levels) in terms of predictors (e.g., the IEUBK model). This is the model we would fit if all variables were observed without error. In what follows, we will call Y the response.
- A variable that is measured subject to error (e.g., exposure to lead via wipe samples). We will call this variable X, e.g., the average environmental lead level one might obtain in wipe sampling if one does many wipe samples per day for a fairly large number of days. In other words, X is the "true" exposure. It is often called the error-prone predictor or the latent predictor.
- The observed value of the mismeasured variable, e.g., the average of a few wipe samples done on a single day. We will call this W.
- Those predictors that for all practical purposes are measured without error (e.g., age, race, gender), which we will call Z.
- We are interested in relating the response Y to the true predictors (Z,X). One method, often called the naive method, simply replaces the error-prone predictor X with its measured version W. This substitution typically leads to biases in parameter estimates and can lead to misleading inferences.
- The goal of measurement error modeling is to obtain nearly unbiased estimates and inferences. Attainment of this goal requires careful analysis. Substituting W for X but making no adjustments in the usual fitting methods for this substitution leads to estimates that are biased, sometimes seriously. In assessing measurement error, careful attention needs to be given to the type and nature of the error and to the sources of data that allow modeling of this error.

Of course, it should be obvious that one should design studies and instruments in such a way as to lessen or eliminate measurement error.

Computer Programs

Splus and SAS computer programs (on SPARC architecture SunOS versions 4 and

5 and for Windows on PCs) are available on the World Wide Web at *http://stat.tamu.edu/qvf/qvf.html*

Models for Measurement Error

There are many models for measurement error (*1*). For purposes of specificity, we will base our discussion on the additive error model, i.e., observed lead exposure W differs from accurately measured lead exposure X because of a random addition of measurement error. The random measurement error will be said to have variance σ_u^2.

We are not suggesting that observed lead exposure (usually, of course, in the log scale) necessarily differs from true lead exposure (again in the log scale) in an additive way. There are many other ways that measurement error can occur. The purpose of this paper is to point out some of the effects of measurement error, and it seems preferable to illustrate these effects in an important special case. Although the formulas and techniques differ depending on the form of measurement error, our three basic points remain essentially invariant to this form.

Carroll et al. (*1*) discuss in detail various ways to understand the form of measurement error.

Transportability of Models and Parameters

In some studies, the measurement error process is not assessed directly, but data from other independent studies (called external data sets) are used instead. We say that parameters of a model can be transported from one study to another if the model holds with the same parameter values in both studies.

In many instances, approximately the same error model holds across different populations. For example, consider wipe sampling at two different locations. Assuming similar levels of training for technicians making the measurements and a similar protocol, it may be reasonable to expect that the distribution of the error in the recorded measure depends only on the long-term result, not on the location, the technician making the measurement, or on the value of X being measured. Thus, in classical error models it is often reasonable to assume that the error distribution is the same across different populations, i.e., transportable.

One of the most common mistakes made in that area is to overdo the idea of transportability; in particular, to transport a correction for measurement error from one study to the next. For instance, although the properties of errors of measurement

may be reasonably transportable, the properties of the true (or latent) predictor X are rarely transportable, as they depend so heavily on the population being sampled, and the corrections for measurement error in the two populations will be strikingly different. As another example, the distribution of true wipe sampling in a population defined in a single area is hardly likely to be transportable to the nation at large. Carroll and Stefanski (*3*) give an explicit example of the dangers of assuming transportability.

Linear Regression and the Effects of Measurement Error

Overview

A comprehensive account of linear measurement error models can be found in Fuller (*2*). Carroll et al. (*1*) give a briefer overview of the essential issues.

In what follows, we will assume for illustrative purposes that blood lead is related to lead exposure linearly (possibly after a logarithmic transformation). We will refer to this as a complex model and will even blur the distinction between this model and the IEUBK model. We hope that the reader will forgive us for these simplifications. Our three main points hold generally, but explicit and easy answers are available in the linear case, and thus are ideal for illustrating the main ideas.

Many textbooks contain a description of measurement error in linear regression, usually focusing on simple linear regression and concluding that the effect of measurement error is to bias the slope estimate in the direction of zero. Bias of this nature is commonly referred to as attenuation or attenuation to the null. We will repeat some of this work but with a more pronounced emphasis on prediction than is typical. However, before proceeding, it is important to place this topic in a broader context.

In general (linear and nonlinear) regression problems, the effects of measurement error can be complex. In multiple linear regression, the effects of measurement error vary depending on: *a*) the regression model, be it simple or multiple regression; *b*) whether the predictor measured with error is univariate or multivariate; and *c*) the presence of bias in the measurement. The effects can range from the simple attenuation described above to situations in which real effects are hidden, observed data exhibit relationships that are not present in the error-free data, and even the signs of estimated coefficients are reversed relative to the case with no measurement error.

The key point is that the measurement error distribution determines the effects of measurement error; thus, appropriate methods for correcting for the effects of measurement error depend on the measurement error distribution.

Simple Linear Regression with Additive Error: Regression to the Mean

We start with the simple linear regression model with intercept β_0, slope β_x and variance about the line σ_ε^2. The true values of the predictor are called X, and with considerable license we will refer to this as true lead exposure, and assume that it has mean μ_x and variance σ_x^2. The error model is additive with error variance σ_u^2.

To illustrate the attenuation associated with the additive measurement error, we simulated data from 10 observations, with $\sigma_x^2 = \sigma_u^2 = 1$, $\beta_0 = 0$, $\beta_x = 1$ and $\sigma_\varepsilon^2 = 25$. In Figure 1, we plot the blood lead levels Y against the true environmental lead exposures X. Note the steep slope in the plot and that the observations are tightly bunched near the line. This indicates that, in actuality, there is a strong and nearly direct relationship between blood lead levels and environmental exposure.

We next illustrate the effects of measurement error by displaying in Figure 2 what might happen if lead exposure were measured with error. In this plot, in addition to the true fits of Figure 1, the empty circles and attenuated line depict the blood lead levels Y and observed, error-prone environmental lead levels W, along with the fitted regression line. There are
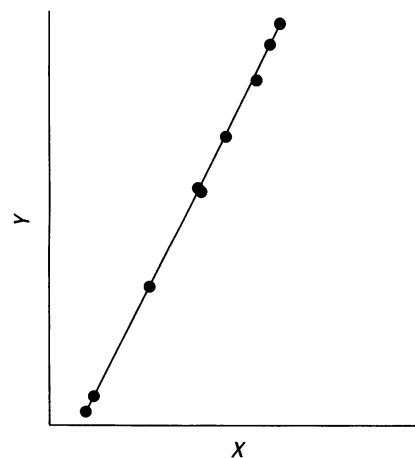

Figure 1. Illustration of additive measurement error model. (●) true (Y,X) data and the line is the least-squares fit to these data.
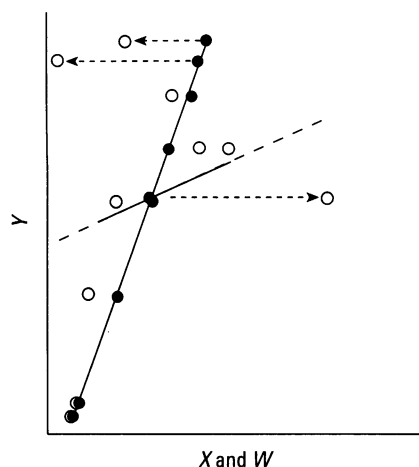
Figure 2. Illustration of additive measurement error model. (●) true (R,X) data and the steeper line is the least squares fit to these data. The empty circles and attenuated line are the observed (R,W) data and the associated least-squares line.

two important, indeed critical, points that this figure illustrates.

a) The effect of ignoring measurement error is to produce a biased estimate of the line. In fact, it is well known that the line fitted with error-prone exposure data estimates not the true slope β_x, but instead $\lambda\beta_x$, where

$$\lambda = \text{reliability ratio} = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} < 1. \quad [1]$$

The attenuating factor, λ, is called the reliability ratio.

b) Figure 2 also illustrates that the fit to the line has seriously degraded. Not only is the line attenuated, but the error about the line has vastly increased. Indeed, while the error about the line with reliably measured lead exposure is σ_ε^2, the error about the line with the error-prone lead exposure measures is

$$\text{residual variance of observed data}$$
$$= \sigma_\varepsilon^2 + \lambda\beta_x^2\sigma_u^2.$$

This facet of the problem is often ignored, but it is important. Measurement error causes a dual difficulty: not only is the slope attenuated, but the data are more noisy, with an increased error about the line.

Figure 2 is indicative of a phenomenon called regression to the mean. Intuitively, what this means is that the extremes in the observed data (in this illustration, lead exposure) are too extreme, and that the true lead exposure is closer to the mean of the data. In fact, in normally distributed data, if true lead exposure has a population mean μ_x, then having observed the fallible instrument, the best prediction of true lead exposure for a single individual with observed lead exposure W is $\mu_x(1-\lambda+\lambda W$, where the reliability ratio is $\lambda < 1$ and is defined in Equation 1. The net effect is that the best (linear) predictor of true lead exposure is always closer to the overall mean than any observed but error-prone lead exposure.

The foregoing is one facet of regression to the mean. A more common definition is complementary. For a child with an extreme observed but error-prone exposure, if one repeats the measurement and obtains a second (replicated) measure, this replicate is generally less, and often much less, than the original extreme value.

Transportability

We now are in a position to see why it is that corrections for measurement error derived from one study should not be applied directly to a second study. The reason is that the reliability ratio (Equation 1) depends critically on the variance of true lead exposure. This variability of lead exposure may differ greatly from study to study, leading to different reliability ratios.

Multiple Regression: Single Covariate Measured with Error

In multiple linear regression the effects of measurement error are more complicated, even for the additive error model. The full details are beyond the scope of this paper; see Carroll et al. (*1*), especially Chapters 2 and 11.

Multiple Covariates Measured with Error

If multiple covariates are measured with error, then the direction of the bias induced by this error does not follow any simple pattern. One may have attenuation, reverse attenuation, changes of sign, an observed positive effect even at a true null model, etc. This is especially the case when the predictors measured with error are correlated or their errors are correlated. With such problems, there really seems to be no substitute for a careful measurement error analysis.

Correcting for Bias

As we have just seen, the ordinary least squares estimator is typically biased under measurement error, and the direction and magnitude of the bias depends on the regression model and the measurement error distribution. The usual method of correcting for such measurement error is the method of moment; see Fuller (*2*), especially Chapter 2.

Another well-publicized method for linear regression in the presence of measurement error is orthogonal regression [see Carroll and Ruppert (*4*) for criticism]. We believe this method is used too frequently.

Prediction

We are now in a position to describe the third of the major points we mentioned in the "Overview." Specifically, it is our contention that if one builds a complex mechanistic model such as IEUBK model using reliable environmental lead exposure data, one can expect that it will do a poor job of prediction when applied to error-prone lead exposures, except possibly in predicting the mean blood lead level.

The point is best made graphically. Consider Figure 3. This is meant to illustrate the fitted prediction line from a complex model built using the best available data. In actuality, the line is $\beta_0 + \beta_x X$, where $\beta_0 = 0, \beta_x = 1$. For a given true lead exposure level X, we predict that on average the blood lead level will be $\beta_0 + \beta_x X$.
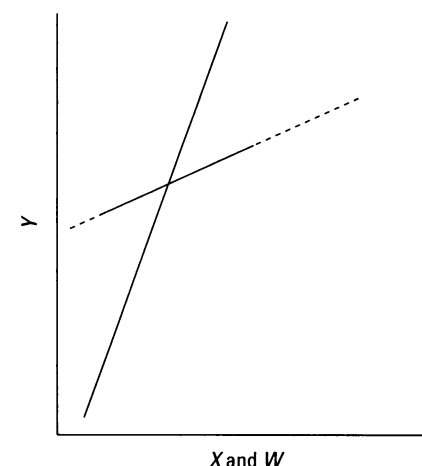

Figure 3. This illustrates the effect of measurement error on prediction. The solid (steeper) line is the prediction line that one would use if there were no measurement error. The dashed line is the predictions one would make if lead exposure were measured with error. Note that while the prediction at the mean observed lead exposure is approximately correct, the predictions are wrong at the high levels of exposure that are typically of interest. In this plot, the mechanistic model (solid line), e.g., IEUBK, will overestimate exceedance probabilities.

In Figure 3, we also add in the (dashed) line that occurs if one has error-prone lead exposure levels. That is, for a given error-prone, observed lead exposure level W, this is the average blood lead level that will be observed. A mathematical justification is given in the "Appendix," but effectively this is the observed (dashed) line in Figure 2 based on large sample sizes.

In considering Figure 3, note what happens. Even though the complex model (e.g., IEUBK model) is a perfectly correct model in relating blood lead levels to true lead exposure, it does a poor job of predicting blood lead levels from error-prone lead exposures. Although the predicted blood lead level at the mean lead exposure is approximately correct, the complex model simply grossly mis-estimates the effect of lead exposure at high levels.

Another way to think of Figure 3 is in terms of exceedances. Suppose that one is interested in the percentage of individuals whose blood lead exceeds a threshold t. That is, one builds a complex model, then applies it to a new data set that has error-prone lead exposures. One method is simply to write down the predictions in the new data and count the percentage of blood lead predictions that exceed the threshold. Figure 2 makes it clear that this prediction

will simply be in error, and thus the true effect of lead exposure on blood lead levels will be misjudged. In the "Appendix," we construct a fictitious situation in which 9% of the population actually exceeds a threshold, but by ignoring measurement error we would estimate that 16% of the population exceeds the threshold.

More complicated procedures for estimating the percentage of a population exceeding a threshold are available; see the "Appendix" for a technical analysis of one such method. The important point is this: If one carefully fits a model such as the IEUBK using reliable data, and then applies this model to error-prone lead exposures, one can expect predictions of the percentage of the population exceeding a blood lead level threshold to have bias, often serious bias. If one really wants to validate a complex model on error-prone exposure data, a more complex process is required that carefully takes into account all facets of the problem, including measurement error. A brief overview of this is given in the "Appendix."

Discussion

We have shown that measurement error of the type one might expect in lead exposure data will bias parameter estimates. Models fit with error-prone exposures will not be accurate indicators of the model that

relates true exposure to blood lead levels. Biases of this type are well known and have been discussed extensively in the statistics literature.

Less well known is the effect of measurement errors on prediction. A model relating blood levels to true lead exposure that is applied to error-prone exposures can be expected to yield a biased estimate of quantities such as the percentage of the population whose blood lead exceeds a given threshold; interestingly, the mean blood lead is typically not badly affected by errors in exposures. In the "Appendix," we construct a fictitious situation in which 9% of the population actually exceeds a threshold, but by ignoring measurement error we would estimate that 16% of the population exceeds the threshold.

What this means is that complex models such as the IEUBK model cannot be validated by applying them to data with error-prone lead exposures. Even if this model is correct in all respects, we have shown that we expect it will not perform very well in estimating probabilities of high blood levels. Although there are statistical approaches to validating the model properly (see the "Appendix" for one such approach at a theoretical level), it remains far easier to validate the complex model on data for which exposure has been relatively carefully ascertained.

Appendix

Estimation of Exceedances

We demonstrate here what might happen when one uses a complex model fit using reliable data and then applies it to error-prone data. We assume that one is interested in estimating the percentage of the population that has a response Y exceeding a threshold t.

We distinguish between the reliable data set and the error-prone data set. We have used the reliable data set to construct a model relating true lead exposure level (X) to blood lead level (Y). In linear regression, this gives good estimates of the intercept (β_0), the slope (β_1), and the variance about the line (σ_ε).

In the error-prone data set, suppose that the true exposures X are normally distributed with mean zero and variance σ_x^2. The assumption that X has mean zero is for notational convenience only, and the conclusions do not depend on this. Generally, we will be interested in thresholds of blood

lead that exceed the intercept of the line, so that the threshold $t > \beta_0$. Then the percentage of the error-prone population that has blood lead exceeding t can be shown to be

$$\mathrm{pr}(Y > t) = E\{\mathrm{pr}(Y > t \mid X)\}$$
$$= E\left\{1 - \Phi\left(\frac{t - \beta_0 - \beta_x X}{\sigma_\varepsilon}\right)\right\}$$
$$= 1 - \Phi\left\{\frac{t - \beta_0}{\left(\sigma_\varepsilon^2 + \beta_x^2 \sigma_x^2\right)^{1/2}}\right\}. \quad [2]$$

Equation 2 is the actual percentage of the error-prone population whose blood lead levels exceed the threshold t.

If we ignore the measurement error in the lead exposure levels, the complex mechanistic model leads us to predict that the following percentage of the population has blood lead levels exceeding t:

$$\mathrm{pr}(Y > t \mid \text{error} - \text{prone})$$
$$= 1 - \Phi\left[\frac{t - \beta_0}{\left\{\sigma_\varepsilon^2 + \beta_x^2\left(\sigma_x^2 + \sigma_u^2\right)\right\}^{1/2}}\right]. \quad [3]$$

As $t > \beta_0$, what we see is that the complex mechanistic model applied to error-prone exposures results in an overestimate of the percentage of the population with high blood lead levels. For instance, if $t = 1.5$, $\beta_0 = 0$, $\beta_x = 1$, $\sigma_x^2 = \sigma_u^2 = 1$, and $\sigma_\varepsilon^2 = 0.25$, then the actual percentage of the population with blood lead levels exceeding the threshold is 9%, whereas the complex model fit using reliable exposure data would predict that 16% of the population exceeds the threshold.

Of course, such overestimation need not always be the case. The IEUBK model is, of course, much more complex than the simple linear model that we have considered, with more than one type of lead

exposure and important demographic characteristics such as gender, age, and race that must be considered. What we can predict in general is that in ignoring measurement error, correct and carefully fit complex models fit to error-prone exposure data will typically do a poor job of estimating the percentage of the population exceeding the threshold.

The Predictive Distribution

Suppose that one has carefully fit a model for Y as a function of X, and has written the density function of this model as $f_{Y|X}(y|x)$. In our context, this model was fit using reliable lead exposure data, and it is assumed to be transportable from this careful study to a second one that has error-prone exposures. In this second data set, the error model is $f_{W|X}(w|x)$. The actual predictive density requires a model for X itself in this second data set, which we write as $f_X(x)$. We assume that the errors in lead exposure measurements are independent of blood lead.

With these assumptions, the density function of blood lead in the second data set is

$$f_y(y) = \int f_{Y|X}(y\,|\,x) f_{W|X}(w\,|\,x) f_x(x) dx dw.$$

The appearance of the error model $f_{W|X}(w|x)$ makes it clear that special and careful attention must be paid to the error process. The appearance of the true exposure distribution $f_X x$ makes it clear that the effects of measurement error differ from study to study, and one cannot simply assume that they are the same across all studies.

REFERENCES AND NOTES

1. Carroll RJ, Ruppert D, Stefanski LA. Measurement Error in Nonlinear Models. London:Chapman and Hall, 1995.
2. Fuller WA. Measurement Error Models. New York:John Wiley and Sons, 1987.
3. Carroll RJ, Stefanski LA. Measurement error, instrumental variables and corrections for attenuation with applications to meta-analyses. Stat Med 13:1265–1282 (1994)
4. Carroll RJ, Ruppert D. The use and misuse of orthogonal regression in measurement error models. Am Stat 50:1–6 (1996).